Prospects for a Naturalistic Explanation of the Good

Christine M. Korsgaard

Abstract: In this paper I explore the possibility of explaining why there is such a thing as the good in naturalistic terms. More specifically, I seek an explanation of the fact that some things are good-for human beings and the other animals in the final sense of good: worth aiming at. I trace the existence of the final good to the existence of conscious agents. I propose that the final good for an animal is her own well-functioning as the kind of creature she is, taken as an end of action, and that having this as her final good makes her better at the activity she is necessarily engaged in, namely living.

I. The Problem

Lately I have been working on a theory of the good which aspires to be, in a certain sense, naturalistic (Korsgaard, 2013; 2014/2015; 2018, chapter 2). The sense of ‘naturalistic’ that I have in mind is explanatory rather than reductive. What I mean is that I am looking for a way to explain why there is such a thing as the good — why some things are good and some are bad, some things are better and some are worse. I seek an explanation that does not appeal to any irreducibly normative facts, and is consistent with a scientific conception of the world, but does not aspire to reduce the good to something else like pleasure or the satisfaction of desire. I accept the sort of argument, associated with G. E. Moore
(1903) but originally advanced by Richard Price (1758) that says that although such things could turn out in fact to be good, or could even turn out to be ‘The Good,’ we cannot simply identify the concept of the good with any natural property or condition. The concept of the good might apply to pleasure or satisfaction, but it is not the concept of pleasure or satisfaction. For reasons I will explain, the sense of ‘good’ that I take to be most fundamental is what I call the ‘final’ sense of good, in contrast with what I call the ‘evaluative’ or ‘functional’ sense of the good. Speaking roughly, something is evaluatively or functionally good when it has the properties that enable it to serve its function or to serve it well. Something is finally good when it is the sort of thing that is worth having, or aiming at, for its own sake.

In this paper, what I have to say will be limited in one important way. Elsewhere I have argued that the notion of something’s being good for someone — that is for some person, animal, or group — is prior to the notion of something’s being good simply or absolutely (Korsgaard, 2013, 2014/2015, section 2). It is prior in the sense that you can construct a conception of what is good absolutely out of a conception of what is good for someone, but not the reverse. If you start with a conception of what is good absolutely, it is impossible to arrive at an intelligible conception of something’s being good for someone, a notion that identifies the ‘for-ness’ relation in the right way. For purposes of this paper I take that as a given, so

---

1 In fact I believe that simply identifying the good with some natural property tends to block any effort to explain why it exists.
the idea I am seeking to explain is that of something’s being good-for someone. So in this paper I will examine the prospects for producing a naturalistic explanation of why there is such a thing as the good, in the sense of what is good-for someone in the final sense of good. In particular, I will explore the prospects of tracing the fact that there is such a thing as the good to the existence of conscious agents, since that is what seems to me to the most hopeful line to explore.

II. A Naturalistic Explanation of why there is Functional or Evaluative Good

I’m going to start by saying a little about the naturalistic credentials of good in the evaluative or functional sense. Something is good in the evaluative or functional sense when it has the properties that enable it to perform its function well. The notion of ‘function’ I am using here is the notion that I take Aristotle to be using when he speaks of a thing’s ‘ergon.’ In ‘Aristotle’s Function Argument,’ I argued that while for some things it makes sense to identify the ‘ergon’ with the thing’s purpose, for more complex entities what Aristotle has in mind is better captured by thinking about ‘how a thing does what it does,’ its way of doing whatever it characteristically does (Korsgaard 2008, pp. 134-140; especially 138-139). One implication of this is that only someone who understands how a thing works really knows what its function is. We all know that a computer serves many purposes: word processing, document storage, providing access to the internet and so on. But it’s the computer scientist who knows how the computer works, and
therefore how it goes about doing those things, who knows its function. To take another example, anyone might say that the purpose of a heart is to pump blood, but it is the cardiologist, who knows how the heart does that, who knows the function of the heart. This concept of function is what makes it plausible for Aristotle to identify the form of a substance with its function (Metaphysics VII.10 1035a17), while at the same time holding that the form of a substance is what the scientific knower grasps and from which she can reason to explain other things about the substance (Metaphysics VII.7 1032a). It is in the sense of ‘how it does what it does’ that we can view a biological organism as having a function, which is roughly speaking, how it goes about living — that is, staying alive and reproducing. Each kind of organism has its own characteristic way of going about those tasks, and in that sense each kind of organism has its own function.

In fact, once we have the distinction between ‘purpose’ and ‘how a thing does what it does’ we can read the distinction back even onto the most simple substances. What is the purpose of a shelf? To store things on. How does it do that? By providing a flat surface to store things on. That second thing is the shelf’s function, the how-it-does-what-it-does. It is what makes a shelf different from other forms of storage like crates and vases, that work by providing appropriately shaped receptacles to store things in, rather than flat surfaces to store them on.²

² I missed this point in ‘Aristotle’s Function Argument.’ See pp. 138-139.
But here a warning is in order. It is tempting to think that ‘how-a-thing-does-what-it-does’ is a more naturalistic notion than ‘purpose.’ This is because it might seem as if there is a plain fact of the matter about how a thing does what it does, while there is no plain fact of the matter about what its purpose is. The purpose is, as it were, in the eyes of the beholder. Certainly, there does seem to be a plain matter of fact, indeed of scientific fact, about how an organism goes about living. If we think that a thing’s function is how-it-does-what-it-does, and that functional goodness means having properties that enable a thing to perform its function well, we may then be tempted to conclude that functional goodness is a natural notion.

But of course naturalism cannot be bought so cheaply. The difficulty is that the ‘what it does’ in the formula ‘how-it-does-what-it-does’ is still in the eye of the beholder. Suppose that, screwing around with some mechanical parts, I create a little device like a wind-up toy that can move forward autonomously on a surface.\(^3\) But I insist that that is not its purpose, and that in fact it has no purpose. Perhaps I constructed it by accident, while engaged in a kind of mechanical doodling.\(^4\) Now

---

\(^3\) Actually there is no fact of the matter about whether it is moving forward or backward. What I mean is that it can move in a straight line. The fact that it is so hard to rid our descriptions of any trace of intentionality is, in a way, what this paper is about. Ideas of functionality and the good are deeply etched into our way of thinking.

\(^4\) We have to say something like that, because otherwise we will have to explain what guided my movements when I was constructing it, and it will be hard to do that without assigning the object a function or a purpose. See *Self-Constitution*, 2.1, and section 4 below.
suppose that, as it turns out, every once in a while, the device tips over and waves its ‘legs’ in the air like a beetle struggling to right itself. Is there a defect in its construction? Only if we assume, as we do in the case of the beetle, that the ‘what it does’ involves continuing to move forward. If the ‘what it does’ is instead is to ‘move forward for a while and then tip over,’ it is acting just as it should. The specification of its function, the ‘how-it-does-what-it-does,’ will differ, depending on whether what it is doing is ‘moving forward’ or ‘moving forward for a while and then tipping over.’ And that’s the problem. The notion of something ‘doing something,’ the notion of agency at its broadest, has a standard of success and failure built right into it. That makes it a normative notion, although I like to think it is a minimal one.

III. The Relation Between Functional and Final Good

This does not mean that our hopes for a naturalistic explanation of why there is such a thing as functional or evaluative goodness are dashed, however. It just means that in order to say what something is doing, we have to have a conception of what its goal is. Even if it has no goal beyond the performance of the activity itself, we have to switch from regarding the activity as a way of doing something to regarding the activity as a kind of goal before we can extract a notion of what is functionally good for the object from its activity. That’s what I did when I raised the question whether what my little invention is doing is ‘moving forward’ or
‘moving forward and then tipping over.’ I was asking which of those two activities is its goal. If its goal is to move forward, then it fails if it tips over and cannot right itself, while if its goal is to move forward for a while and then tip over, it is doing just fine. In other cases, where the activity is the means to achieving some goal outside of itself, it is even more obvious that we cannot understand what something is doing unless we have a view of what it is supposed to achieve or is trying to achieve by doing that.

This is one reason why the movements of rivers, tectonic plates, forest fires, and erupting volcanoes do not count as things they do in the sense that gives rise to standards of functional goodness.\(^5\) We do not say of such things that they are functioning well or badly. Or at least, we don’t unless we view them as performing some sort of an ecological service. We might think of the forest fire as burning off the brush so that the forest can renew itself, or the river as providing habitat for a certain kind of marine life. Then a fire that burns out too quickly or a sluggish river might be doing badly. This is all just a way of belaboring the obvious: the notion of

---

\(^5\) Some readers will demur about some of these cases, for two reasons. The first is that we are tempted to think that rivers are flowing well and fires are burning well when they are flowing or burning in a way that makes them likely to keep flowing or burning. This is because we transfer a thought that is natural to have about living things to these non-living phenomena. Living things are living well when they are living in a way that makes it likely that they will keep living. The other reason is even more primitive. People may be tempted to think a fire is burning well or a volcano is erupting well when they are doing whatever they do vigorously. If I said a volcano was erupting feebly, someone might think I was criticizing it. I will not venture an explanation of that one.
functional or evaluative good depends upon a notion of final good, in the sense of something regarded as a goal. So then the question becomes whether we can give a naturalistic explanation of why there is such a thing as final good.

IV. A Naturalistic Explanation of why there is Final Good?

Can we give a naturalistic explanation of why there is final good? If we just take something’s final good to be a goal that is there is some reason to attribute to it, this seems easy. There is final good because there are conscious agents, who really are trying to do certain things and to achieve certain things. This suggests that final goods exist in the perspective of conscious agents. It is a perspectival notion.

This in itself does not detract from the naturalistic credentials of the idea, because we are not trying to reduce the notion of the final good to something else like pleasure or the satisfaction of desire; we are only trying to explain why it exists without appealing to irreducibly normative facts. A perspectival notion can have perfectly respectable naturalistic credentials in this sense. For example, I suppose that the natural sciences do not, or do not need to, traffic in furniture and artifact concepts. Furniture and artifact concepts are functional concepts with an evaluative or normative dimension. To know that something is a chair is to know what it is for, and that already is to know something about the evaluative or functional standards that apply to it. But for purposes of scientific explanation, we do not usually need to refer to the fact that a certain object is a chair. Much of the
time, it is enough for the purposes of scientific explanation that it is an object of a certain shape, size, and composition. On certain occasions, though, especially when we are giving psychological explanations, it might be necessary to refer to the fact that someone thought it was a chair. That is no problem, for there is a perfectly good scientific explanation of why there are creatures who think of some things as chairs, who conceptualize the world in that way. It is because there are creatures who conceptualize the world, and some of them are bipedal creatures who sometimes need to sit down in the way made possible by chairs. All we need to do is produce a scientific explanation of how such creatures came to be in the world — that is, an account of human evolution. Although it is only in the eyes of someone who knows what it means to sit down that anything is a chair, as long as we can explain the existence of the perspective within which the concept of a chair appears, there is nothing non-natural or spooky about chairs. Unless, of course, there is something non-natural or spooky about thought.

The situation with the final good is similar. In ‘Realism and Constructivism in Twentieth-Century Moral Philosophy,’ I argued that self-conscious agents, agents who know what they are doing, operate with normative concepts because we have normative problems to solve (Korsgaard 2003/2008, section 6). In other words, we should give a what I have called a ‘constructivist’ account of why there is such a thing as the good. According to a constructivist account of our basic normative concepts, we use the concept of the final good because we have to decide what to
aim at, and we use the concept of the right because we have to decide what to do. We face these problems, and we use those concepts when we need a kind of placeholder word for whatever it is that solves those problems, or when we want to conceptualize or announce the fact that we have solved them. We ask what would be best, as a way of asking what would solve the problem of what to aim at. We say things like ‘oh, that will be good!’ when we think we have solved it (see also Korsgaard, 1996, section 3.2.1). So to explain why there are creatures who must conceptualize the world using the concept of the final good, what we have to do is explain how there came to be creatures in the world who must act, and who therefore face the problem of deciding or determining what to do. There is a natural explanation of that, and we all know roughly what it is. It is that conscious living animals, who are agents with various degrees of cognitive sophistication, evolved on this planet. Some of those agents, namely rational agents, know what we are doing in part by conceptualizing what we are doing, and in order to do so, we must use the concept of a final good or a goal.

But here we run into a complication. We do not want to say that when something is perspectival in this way, it exists only in the perspective of creatures who can conceptualize it, or only in the perspective of creatures who can conceptualize it in the linguistic way that we do. Furniture and artifacts do not exist only for rational beings. If you put up a birdhouse in your back yard in hopes of attracting a bluebird family, and a pair of bluebirds come and check it out, it is...
clear that there is some sense in which they perceive it as a potential nesting place.
These are animals who need nesting places just as we need places to sit down, and, like us, they view the world through the eyes of their needs. When they build their own nests, their movements are guided by some sense of how the nest is to be used. The exact sense in which it is correct to say that an animal sees something as a possible home — the form that that perception takes in his or her mental life — probably varies with the degree and kind of cognitive sophistication of the animal in question. So when we say that a bluebird sees a birdhouse as a possible nesting place, and when we say that a termite sees your house as a possible nesting place, we may be talking about rather different things, different things going on in the mental lives of these animals. But they are things that serve a similar function — a function that is similar to the function that the concept of a home plays in our mental lives.

The reason this matters is that we don’t want to say that things are only finally good or bad in the perspective of rational or conceptualizing agents. Final good and bad exist in the perspective of any agent who is trying to achieve something in particular and has some idea what he or she is doing.\(^6\) Obviously, the idea applies most clearly to rational agents, who say to ourselves that such-and-

\(^6\) This is a bit rough, but I am trying to use ‘concept’ for a kind of certain occurrence in the mental life of a rational/linguistic animal, and ‘idea’ for whatever it is in the life of any animal that plays the same functional role as a certain concept.
such is what we are trying to achieve, and such-and-such is the way to go about it. Some people would no doubt like to draw a big hard line there, and say that animals who cannot do that (all the others?) are driven by instinct and do not know what they are doing. But I think that the nature of action makes that implausible, for a reason I will explain in a moment.

I think what we have is a more of a continuum or a hierarchy. At the lower end, there are creatures whose movements we think of as having an evolutionary purpose — as ‘designed’ by natural selection to facilitate their survival and reproduction. That evolutionary purpose may not appear in the animal’s mental life except as the urge to respond in a certain way to a certain stimulus. At the higher end, there are rational agents, who think about what we are trying to achieve and reason instrumentally about how to go about it. In between there are agents who must have some idea what they are trying to do or even what they are trying to achieve by it, because, being agents, they monitor their own movements and adjust them constantly in ways that make it more likely that they will succeed. This is the point about agency that makes it implausible to think that no non-rational animals have any idea what they are doing. You only have to think for a minute to realize just how far down the hierarchy I just described this has to go. Almost any agent who can locomote will go around obstacles rather than just stopping in front of them. Perhaps there are some organisms so primitive mentally that what happens next has to be determined by a new impulse, but for many what happens next is
that they get back on track of what they were doing. But on track of what, if they have no idea of what they are doing? It is impossible to suppose that inborn instinct is sufficiently refined to provide for every eventuality an acting animal has to deal with as she acts. But to what extent an animal knows what she is doing is a matter of degree. Suppose a fish’s instincts tell it to snap at a disturbance on the surface of the lake. In the world in which the fish evolved, such disturbances are nearly always caused by edible insects. This particular disturbance is caused by a piece of litter that some human slob just threw into the water. It is not something to eat, so the fish spits it out. The really dumb fish may just snap at it again as it moves along the surface of the water. The smarter fish moves on. The smarter fish then has some idea, however dim, that he is looking for something to eat, and the litter didn’t answer to that idea. On up from there to lions who coordinate their movements so that they are approaching the wildebeest from every side, or to Sultan the chimpanzee who stacks up crates to reach the high-hanging bananas, to Alex the language-trained parrot who said ‘I want a nut’ when he wanted a nut (Köhler, 1924/2013; Pepperberg 2000). What we have here is a matter of degree. Even if animals do not use the concept of a final good or a goal, they do something that to varying degrees is functionally like being guided by the concept of a final good or a goal and how to achieve it. And recognizing that, we can use the concept of a final good or a goal on their behalf.
V. Giving Content to the Final Good

There are some problems with this line of argument, which I will come back to, but for now I will say that it looks as if we can give a naturalistic explanation of why we must use the concept of a final good or goal, and even of why we can apply it to other agents, those of the other animals who count as agents, as well as to ourselves. The tricky bit is explaining why this concept is one that can be applied correctly or incorrectly. For we have not explained why there is such a thing as a final good unless we have explained not only why we use the concept, but why it is a concept that can be applied correctly or incorrectly — why it actually picks something out something real and determinate. To put the same point another way, there are only final goods if some things are actually worth pursuing, at least, to the animal who pursues them. Although this is controversial, I believe that this thought is built right into the standard of success and failure, that, as I argued earlier, is part of the very idea of action. Failure is not merely the failure to do whatever an agent was trying to do, but the failure to do something that, at least from the agent’s point of view, is actually worth doing, whether the agent was trying to do that or not.

---

7 I am using ‘agent’ here to refer to any animal that uses locomotion guided by perception to satisfy some of its needs, for food, safety, to locate mates, etc. Stationery filter-feeding animals like sponges may not count.

8 It is controversial because what I am about to say is a defense of the thesis that we act under ‘the guise of the good.’
Here’s why: earlier I argued that even cognitively unsophisticated agents must have some idea what they are doing, because they monitor their movements in ways that are designed to stay on track of what they are doing. They go around obstacles, for example. But along with that — really, part and parcel of that — is the fact that there is some measure of how much pain, effort, or risk an agent is willing to endure before she desists from trying to stay on track of what she is doing. The hungry fox will chase the rabbit into the brush, say, even if that means getting scratched by thorns, but not into a fire or into close proximity with humans or wolves. The fact that there is such a measure suggests that the agent is in some way guided by an idea of what the end is worth to her. But she cannot have an idea of what the end is worth to her unless it is worth something to her. So a final good is not just a goal, but a goal that is conceived to be worth pursuing. If that is true, there are only final goods if some things are worth pursuing.

Many philosophers are prepared to tell us what the concept of the final good picks out, that is, to tell us what the final good is. Current favorites include pleasure, the satisfaction of desire, the satisfaction of well-informed desire, or an ‘objective list’ of activities, experiences, conditions, and what have you that are supposed to be ‘intrinsically’ valuable. What is less clear is why these things are

---

9 As I argued in ‘The Myth of Egoism,’ the idea of ‘the satisfaction of desire’ is ambiguous, since it can refer to the subjective satisfaction you take in getting what you want or the objective state of affairs that you wanted to obtain. On the second view someone’s good is the set of the objects of
supposed to be candidates for the final good. Moore, whose theory is a version of the objective list view, proposes that we use intuition (Moore, 1903, 1922). Supporters of the other views have been accused of doing so as well. To the extent that the accusation is justified, I believe the problem here is that supporters of these theories tend to suppose that these things are worth pursuing because they are good. Indeed, the objective list theorist seems to be pretty much stuck with that conclusion.

On a constructivist account, the reverse should be true: we should deem things to be good because they are worth pursuing, rather than being worth pursuing because they are good. Constructivist accounts generally have this structure. Consider some comparisons. T. M. Scanlon thinks actions are not unjustifiable because they are wrong, but rather wrong because they are unjustifiable (Scanlon, 1998). The kind of relationship which Scanlon thinks we have reason to want to stand in with others demands that we act in ways that we can justify to them. We can use the words ‘right’ and ‘wrong’ as placeholders for whatever it is enables us to do that or not to do it, or when we want to conceptualize or announce someone’s success or failure in acting justifiably.

Immanuel Kant thinks actions are not contrary to law because they are wrong, but wrong because they are contrary to law — that is, they cannot be universalized desire he or she most wants. Since the view that subjective satisfaction is the good is implausible, I am taking this sort of theory in the second sense.
(Kant, 1785/1998). As autonomous beings we must govern our conduct by laws we make for ourselves, and we use the words ‘right’ and ‘wrong’ as placeholders when we ask how to do that, or to announce success and failure in doing it. In a similar way, we have the concept of the good because we have to figure out what to pursue or aim for, and part of that is figuring out what is worth pursuing or aiming for. Since that is the problem, the good should be whatever solves that problem, and the problem is solved when we find something worth pursuing.

You may think that ‘good’ and ‘worth pursuing’ are ideas so close to each other that we have no hope of using one of them to determine the content of the other. How do we determine what is worth pursuing, if it is not by first determining what is good? Looked at from this point of view, the more ‘naturalistic’ of the extant theories of final good, the ones that identify the good with pleasure or the object of desire, do appear to have at least made it to the starting point. Things that are pleasant or desirable at least appear to be worth pursuing to those who experience or expect to experience the pleasure or to those who have the relevant desires. They beckon; they summon an agent to action. Indeed, it may be argued

---

10 On the other hand, you might think they are quite different. An objection that I do not discuss in this paper, because of space constraints, is that not all final goods can be aimed at. Happiness is an example; I am sympathetic to the view that you do not achieve happiness by aiming at it, but rather by trying to do something worthwhile and succeeding. I think it is enough for my purposes that this kind of good is the upshot of action, but I will not try to make the case here.
that that’s what it is to desire something or find it pleasant: for it to look worth pursuing, and \textit{in that sense} to look good. As Aristotle puts it:

To perceive then is like bare asserting or knowing; but when the object is pleasant or painful, the soul makes a quasi-affirmation or negation, and pursues or avoids the object. To feel pleasure or pain is to act with the sensitive mean towards what is good or bad as such.

\textit{(On the Soul III.7 431a7-11)}

So perhaps all we need to ask is whether there is any reason to believe these appearances are true, or to put the point more generally although more vaguely, whether we have some other ground for endorsing them.

I believe that there is, but I cannot even begin to defend the answer here. All I can do is sketch it, so you must regard what I am about to say as nothing but a promissory note.\textsuperscript{11} I think the standard of correctness comes from the idea of functional good. If — admittedly, a big \textit{if} — we regard animals as having the function of living — that is, of surviving and reproducing in the way that is characteristic of their kind, then animals do have a functional good in the fairly straightforward way that I mentioned earlier: there is a fact of the matter, even a scientific fact of the matter, about how they go about living, and therefore a fact of the matter about what counts as being good at it. But animals are ‘designed’ by the

\textsuperscript{11} See ‘On Having a Good,’ and chapter two of \textit{Fellow Creatures: Our Obligations to the Other Animals}, for a fuller defense of the view.
process of evolution to find the things that are functionally good for them attractive, and to find the things that are functionally bad for them aversive. In other words, animals are designed in such a way that they tend to take the things that are functionally good for them as final goods, in the sense that they find them attractive and aim for them, and to reject the things that are functionally bad for them as bad in the final sense in the sense that they find them aversive and try to avoid them. These appearances, however, are highly defeasible, and can be correct or incorrect. So something that appears worth pursuing is only actually worth pursuing if pursuing and obtaining it really does contribute to the functional success of the animal. Actions and ends are good for her if they make the animal good at living in the way characteristic of her kind.

Now, you will want to object that, as I have myself already pointed out, the judgment that what enables an animal to live a life of her kind is good for her in the final sense only makes sense if we regard living a life of her kind as a goal for her, and indeed as a goal worth pursuing, at least from her own point of view. To defend that claim involves, in my view, making two arguments. The first would aim to show that animals themselves take living in the way characteristic of their kind as a goal, by showing that when they monitor their own actions, try to stay on track of what they are doing, and decide when it is or is not worth trying to stay on track,

---

12 The appearances are defeasible for a specific reason: animals that evolve to be fit for one set of circumstances may end up living in different circumstances.
those decisions are guided by the ends of survival and successful reproduction.

That is what it means for something to appear in an animal’s mental life as her goal. That would explain why, for instance, the fox desists from pursuing the rabbit when doing so brings him too near to the humans or the wolves.

I would like to argue here that we know that something along these lines is roughly true of people. When people learn that something that attracts them is likely to kill them, they desist from pursuing it or at least think that they ought to.

But I am hampered here by (at least) two things. First, I would have to convince you that is plausible to say that the human good is simply living in the way characteristic of our kind. I believe this is true, but it only becomes plausible when we grasp some important facts about what the human way of living involves, and that takes a long argument. Living the life of our kind involves maintaining a what I have elsewhere called a practical identity, and that means, speaking roughly, that we are subject to moral as well as physical death: we can feel to meet the standards that, from our own point of view, make our lives worth living and our actions worth undertaking. The second is that I would have to make a case against clear-eyed akrasia, so that you cannot simply reply that it is not true that we avoid things that we know are bad for us.

---

13 For the notion of practical identity see The Sources of Normativity, especially Lecture 3, and Self-Constitution, 1.4, pp. 18-26.
The more complicated argument is to show that there is something right about an agent’s pursuing the goal of living, or some other reason to endorse it. One possibility would be to argue that there is no point of view from which the agent could possibly be wrong. Agency places us in the normative realm, and once we are there, there is no way to refute the claim that something that we take to be good for us is not actually good for us except by showing that it is bad for us. Another possibility, compatible with that, would be to recognize that we rational beings, who are articulate about all this, endorse the goal of living a good life (that is, of living our sort of life well), and we have no more reason to do so than any other animal. Such an endorsement is simply our way of being alive. Life takes itself as a goal, and that is no accident: it is part of what it means to be alive (Korsgaard, 2018, section 8.6).

I realize that this raises many questions. I hope to answer them elsewhere. For purposes of this paper, I’d like to stick to the metaphysical question whether, if we could make such an account plausible, it would give us a naturalistic explanation of the good, in the sense of something’s being good for someone in the final sense of good.

VI. Constructivism and the Problem of Valence

There is (at least) one more problem. I am proposing that we use the concept of the good, in the final sense of good, to indicate the solution to a
problem. The problem is the problem of what to aim at. But the very idea of a problem and a solution already seems to import a kind of normative thought. Why is it a problem that you don’t know what to aim at? It wouldn’t be a problem if anything you aimed at was ‘just as good’ as anything else, would it? You could just pick something. And if it weren’t a problem, of course it wouldn’t have a solution either. So if we suppose that having to decide what to aim at presents us with a problem that needs a solution, aren’t we already supposing that some things are good and others are not, or at least that some things better than others? And if some things are better than others, doesn’t that have to be because some things are good? How then can we appeal to the fact that we have to decide what to aim at to explain why there is such a thing as the good?

Of course the way I just put that isn’t quite right. It is not true that if nothing were better than anything else, you could just pick something, because you wouldn’t have an array of options to choose from, or rather to pick among. Something has to make the different things that we might aim at eligible to pursue in the first place in order to determine what the array of options is. As I said

14 It’s not clear what ‘better than’ means. In some contexts it seems to mean ‘closer to the good,’ (‘He’s been sick, but he’s getting better’) in others something more like ‘includes more good’ or ‘has more good in it’ (‘The vanilla is good, but the chocolate is even better’). But it has to have something to do with the good.

15 The parallel point for theoretical reason is that some things have to be salient for us before we can start to form beliefs.
before, the strength of the more ‘naturalistic’ theories of the good is that they start
from the fact that certain things simply present themselves to us as eligible to be
pursued.

But how do they present themselves that way? This, I think, is where we run
into the most serious challenge for a naturalistic explanation of the good, which I
will call the problem of valence. The problem is that an explanation of why there is
such a thing as the good is almost certainly going to appeal to the fact that some
things have a positive valence, and others have a negative valence, and it is hard to
know how to explain that, or even to say what it means, without appealing to the
idea of the good. Consider the hedonistic theory that pleasure is the good. What
exactly is pleasure, and how does it make things appear worthy of pursuit? There
are a variety of theories about what it is. Bentham appeared to think that it is a
single sensation, varying only in intensity and duration. But what is it about that
sensation that makes agents want to attain it, if not that it seems — well, good?
Hume thought that the word names a type of sensation. He says ‘...under the term
pleasure, we comprehend sensations, which are very different from each other, and
which have only such a distant resemblance, as is requisite to make them be
express’d by the same abstract term.’ (Hume, Part I, Book 3, Section 2, p. 472). But
in what respect do they resemble each other, and how do they differ from pains,
except that the pleasures seem to the agents to be good experiences and the pains
bad ones? Sidgwick thought that pleasure is agreeable consciousness. But what is it to be agreeable, if it is not to seem good?

Of course, hedonists don’t propose to solve the problem of why there is such a thing as the good. Some of them think that pleasure just is the good, and pain just is the bad, and it is a little unclear why we should think so or how we know. But I believe that what makes the theory so perennially tempting is valence. Pleasure and pain just seem to be essentially valenced experiences — in fact it seems as if their whole nature is given by their valence. Pleasure just is positive experience and pain just is negative experience. If you think that the capacity for a final good must have something to do with the character of experience, it is tempting to locate the good in an essentially positively valenced experience, and that is pleasure.

But even if hedonism is false, almost any account of why there is such a thing as the good is going to appeal to the idea that some things are positively valenced. As an example, consider T. M. Scanlon’s ‘buck-passing’ account of the good (Scanlon 1998, chapter 2). A buck-passer is someone who believes that to say that something is good is just to say that there are reasons to have certain attitudes towards it and treat it in certain ways. The fact that these reasons are of different kinds explains the fact that there are different kinds of value, calling for different responses from us. There are some things that there are reasons to promote, some things there are reasons to respect, some things there are reasons to admire, and so
on. On this theory, the existence of the good can be explained, in terms of the existence of reasons. That is not a naturalistic explanation, perhaps, but it is an explanation. But there is still a problem, having to do with the distinction between good and bad. After all, a buck-passer about the good will also be a buck-passer about the bad: the bad too will involve having reasons to have certain attitudes towards things and treat them in certain ways. And since these reasons are also various in nature, we might wonder how the buck-passer is going to distinguish the good from the bad. To say either that something is good or that it is bad is to say that it generates a bunch of practical reasons, so how are the two ideas different? What accounts for the positive valence of the good and the negative valence of the bad? In Scanlon’s account, this problem appears to be passed on to the attitudes and actions for which there are reasons: to claim that something is good is to say there is reason to have positive attitudes and intentions towards it, while to say that something is bad is to say that there are reasons for negative attitudes and intentions. The attitudes and intentions are valenced. But the fact of valence itself remains unexplained.

Since we are not looking for a reductive account of the good, it is not a problem if we explain the existence of the good in terms of the fact that some things have a positive valence. But to prevent circularity we must then have an account of why things have positive or negative valences that does not in turn appeal to the existence of the good.
The solution to this problem, I believe, is, once again, to explain valence in terms of conscious action. Earlier I claimed that the good for an animal is the animal’s functional good — the things that enable the animal to successfully live a life of her kind. Animals are ‘designed’ by the process of evolution to find the things that are functionally good for them attractive, and to find the things that are functionally bad for them aversive. In other words, animals are designed in such a way that they tend to take the things that are functionally good for them as final goods. The trick is to find a way to explain the fact that animals find certain things attractive or aversive without appealing to the idea that those things just seem good or bad to the animals themselves. Or rather, without saying that they seem pleasant or painful, where all that amounts to is a way of seeming good or bad.

Instead, I believe what we should say is that when an animal finds something attractive and is inclined to go for it, what is happening is that she is in fact experiencing herself going for it. Her pleasures and pains are her awareness of the operations of her impulses, or rather — to put it a better way — they are the animal’s having these impulses consciously. The philosopher who comes closest to saying something like this is Hobbes. In his book on Human Nature, Hobbes characterized pleasure and pain as the internal beginnings of animal motion towards or away from the things that help or hinder what he called the animal’s vital motions, or life processes (Hobbes 1650/1991 chapter VII). Hobbes didn’t talk about consciousness in this context, but what he seems to mean is that pleasure
and pain are the animal’s awareness of those beginnings of internal motion. On this view, pleasure and pain are not valenced experiences. They are experiences of valence. Positive valence is just the impulse to go towards something or try to realize it or make it keep going; negative valence is just the impulse to get away from something or try to prevent it or make it stop. Pleasure and pain are our consciousness of these tendencies: or rather, as I said before it is having these tendencies consciously. It is the action that is valenced, as a kind of going towards or away from, realizing or eliminating. These ideas are related to the ancient identification of the real and the good; on this picture, pleasure is movement towards being; pain is movement away from it. But these valences of action are experienced by animals who know what they are doing, or rather to the extent that they know what they are doing.

I realize one worry here is that this makes it sound as if pleasure and pain, or attraction and aversion, are epiphenomenal. The animal is already, as it were, starting to act, so what is their role? But here it is important to recall the self-monitoring that is a feature of all action. Presumably, having the tendency to pursue something consciously enables the animal to exercise that tendency in a way that is more refined and responsive to the environment. On this view, an animal’s experiencing attraction and aversion is experiencing herself trying to get or avoid something. It is a way of knowing what she is doing, and an animal who knows what she is doing will be better at doing it.
If something along these lines is right, we do not have to appeal to essentially valenced experiences in order to explain why there is such a thing as the good. We only need to appeal to the existence of conscious agents, who have some idea what they are doing. Attractions and aversion are just an animal’s consciousness of her own attempts to live in the way characteristic of her kind, to realize her own sort of being. Having that consciousness makes her better at making those attempts. In our own case, that consciousness becomes articulate, which is why we face the problem of having to decide what to aim at and how.

VII. Conclusion

In this paper my aim has been to explore the prospects for giving a naturalistic explanation of why there is such a thing as the good, in the final sense of something’s being good-for-someone. I have tried to sketch an explanation of the existence of the good in terms of the existence of conscious agency. Conscious agents, animals, are doing something already, namely living, and having a final good enables them to do that more effectively. Although it sounds odd to put it this way, having a final good is adaptive. I am painfully aware that my account has left many unanswered questions and that there are many places where it can be challenged. A great deal of work would have to be done to make this story plausible. But I think that there must be an explanation of why there is such a thing as the good, and since it is conscious agents who have a good, in the final sense of
good, it seems to me that the best prospects for explaining the existence of the good is to appeal to requirements of conscious agency itself.

References


Korsgaard, Christine M. *Fellow Creatures: Our Obligations to the Other Animals*. Forthcoming from Oxford University Press, 2018.


---

Christine M. Korsgaard  
Arthur Kingsley Porter Professor of Philosophy  
Harvard University  
Department of Philosophy  
205 Emerson Hall  
Cambridge, MA. 02138  
USA  
christine_korsgaard@harvard.edu